AMENDMENTS TO THE CLAIMS:

This listing of claims will replace all prior versions and listings of claims in the application:

Listing of Claims:

1. **(currently amended)** A method of clustering documents ~~or patterns~~ each having one or plural document ~~or pattern~~ segments in an input document ~~or pattern~~ set, said method comprising:

(a) obtaining a co-occurrence matrix for each input document, and obtaining an input document ~~or pattern~~ frequency matrix for the set of input documents ~~or patterns,~~ based on occurrence frequencies of terms or term pairs appearing in the set of input documents ~~each document or pattern~~;

(b) selecting a seed document ~~or pattern~~ from a set of remaining documents ~~or patterns~~ that are not included in any cluster existing at that moment, and constructing a current cluster of an initial state based on ~~using~~ the seed document ~~or pattern~~, wherein said selecting and constructing comprise; ~~comprises~~

(b-1) constructing a remaining document common co-occurrence matrix for the set of the remaining documents based on a product of corresponding components of the co-occurrence matrices of all documents in the set of remaining documents; ~~or patterns; and~~

(b-2) obtaining a document commonality of each remaining document to the set of the remaining documents based on a product sum between every component of the co-occurrence matrix of each remaining document and the corresponding component of the remaining document common co-occurrence matrix;

(b-3) ~~(b-2) using the common co-occurrence matrix to~~ extracting, as the seed document ~~or pattern~~, the document ~~or pattern~~ having the highest

document ~~or pattern~~ commonality to the set of the remaining documents ~~or patterns~~; and

(b-4) constructing the initial cluster by including the seed document and neighbor documents similar to the seed document;

(c) ~~obtaining the document or pattern commonality to the current cluster for each document or pattern in the input document or pattern set by using information based on the document or pattern frequency matrix for the input document or pattern set, information based on the document or pattern frequency matrix for documents or patterns in the current cluster and information based on a common co-occurrence matrix of the current cluster, and~~ making documents, which have ~~or patterns having~~ the document commonality to the current cluster higher than a threshold, belong temporarily to the current cluster; wherein said making comprising:

(c-1) constructing a current cluster common co-occurrence matrix for the current cluster and a current cluster document frequency matrix of the current cluster based on occurrence frequencies of terms or term pairs appearing in the documents of the current cluster;

(c-2) obtaining a distinctiveness value of each term and each term pair for the current cluster by comparing the input document frequency matrix with the current cluster document frequency matrix;

(c-3) obtaining weights of each term and each term pair from their distinctiveness values;

(c-4) obtaining a document commonality to the current cluster for each document in the input document set based on a product sum between every component of the co-occurrence matrix of the input document and the corresponding component of the current cluster common co-occurrence matrix while applying the respective weights to said components; and

(c-5) making documents having the document commonality to the current cluster higher than the threshold belong temporarily to the current cluster;

(d) repeating step (c) until the number of documents ~~or patterns~~ temporarily belonging to the current cluster ~~becomes the same as that in the previous repetition~~ does not increase;

(e) repeating steps (b) through (d) until a given convergence condition is satisfied; and

(f) deciding, on the basis of the document ~~or pattern~~ commonality of each document ~~or pattern~~ to each cluster, a cluster to which each document ~~or pattern~~ belongs and outputting said cluster.

2. **(currently amended)** A clustering method according to claim 1, wherein said step (a) further includes:

(a-1) generating an input document ~~or pattern~~ segment vector for each of said input document ~~or pattern~~ segments based on occurrence frequencies of terms appearing in each input document ~~or pattern~~ segment;

(a-2) obtaining a co-occurrence matrix for each input document from ~~or pattern in~~ the input document ~~or pattern set from the document or pattern~~ segment vectors; and

(a-3) obtaining an input document ~~or pattern~~ frequency matrix from the co-occurrence matrix for each document.

3-4. (canceled)

5. **(currently amended)** A clustering method according to claim 1, wherein ~~further including:~~

the convergence condition in said ~~repeating~~ step (e) is satisfied when (i) ~~until~~ the number of documents ~~or patterns~~ whose document ~~or pattern~~ commonalities to any current clusters are less than a threshold becomes 0, or (ii) the number is less than a threshold and ~~is equal to that of the previous repetition~~ does not increase.

6. **(currently amended)** A clustering method according to claim 1, wherein said step (f) further includes:

checking existence of a redundant cluster, and removing, when the redundant cluster exists, the redundant cluster and again deciding the cluster to which each document belongs.

7. **(currently amended)** A method of clustering documents ~~or patterns~~ each having one or plural document ~~or pattern~~ segments in an input document ~~or pattern~~ set, said method comprising:

(a) obtaining a co-occurrence matrix S^r for each input document D_r ~~a document or pattern frequency matrix for the set of input documents or patterns,~~ based on occurrence frequencies of terms or term pairs appearing in ~~each document or pattern~~ the set of input documents;

(b) selecting a seed document ~~or pattern~~ from a set of remaining documents ~~or patterns~~ that are not included in any cluster existing at that moment, and constructing a current cluster of an initial state ~~using~~ based on the seed document, wherein said selecting and constructing comprise:

(b-1) constructing a remaining document common co-occurrence matrix T^A for the set of the remaining documents based on the co-occurrence matrices of all documents in the set of remaining documents;

(b-2) obtaining a document commonality of each remaining document to the set of the remaining documents based on the co-occurrence matrix S^r of each remaining document and the remaining document common co-occurrence matrix T^A;

(b-3) extracting, as the seed document, the document having the highest document commonality to the set of the remaining documents; and

(b-4) constructing the initial cluster by including the seed document and neighbor documents similar to the seed document; ~~or pattern;~~

(c) ~~obtaining the document or pattern commonality to the current cluster for each document or pattern in the input document or pattern set by using information based on the document or pattern frequency matrix for the input document or pattern set, information based on the document or pattern frequency matrix for documents or patterns in the current cluster and information based on a common co-occurrence matrix of the current cluster, and~~ making

documents ~~or patterns~~ having the document commonality higher than a threshold belong temporarily to the current cluster;

(d) repeating step (c) until the number of documents ~~or patterns~~ temporarily belonging to the current cluster ~~becomes the same as that in the previous repetition~~does not increase;

(e) repeating steps (b) through (d) until a given convergence condition is satisfied; and

(f) deciding, on the basis of the document ~~or pattern~~ commonality of each document ~~or pattern~~ to each cluster, a cluster to which each document ~~or pattern~~ belongs and outputting said cluster;

~~wherein a~~

wherein in step (a), each mn component S^r_{mn} of the co-occurrence matrix S^r of the document ~~or pattern~~ D_r is determined in accordance with:

~~$$S^r = \sum_{y=1}^{Y_r} d_{ry} d_{ry}^T \quad (1)$$~~

$$S^r{}_{mn} = \sum_{y=1}^{Y_r} d_{rym} d_{ryn}$$

where:

m and n denote m^{th} and n^{th} terms, respectively, among M ~~is the number of sorts of the occurring~~ terms appearing in the set of input documents,

D_r is the r^{th} document ~~or pattern~~ in a document ~~or pattern~~ set D consisting of R documents ~~or patterns~~,

Y_r is the number of document ~~or pattern~~ segments in document ~~or pattern~~ D_r, and

~~$d_{ry} = (d_{ry1},.., d_{ryM})^T$ is the y^{th} document or pattern segment vector of document or pattern D_r, and T represents transposition of a vector~~ wherein d_{rym} and d_{ryn} denote the existence or absence of the m^{th} and n^{th} terms, respectively, in the y^{th} document segment of document D_r, and

S^r_{mm} represents the number of document segments in which the m^{th} term occurs and S^r_{mn} represents the co-occurrence counts of document segments in which the m^{th} and n^{th} terms co-occur.

8. (canceled)

9. **(currently amended)** A method according to claim 7, ~~further comprising:~~

wherein in step (b-1), ~~determining~~ the remaining document common co-occurrence matrix T^A ~~of the document or pattern set D~~ is determined on the basis of a matrix T;

wherein

~~an mn component of S^r is given by~~

~~$$S^r_{mn} = \sum_{y=1}^{Y_r} d_{rym} d_{ryn}$$~~

the matrix T has an mn component determined by

$$T_{mn} = \prod_{r=1, S^r_{mm}>0}^{R} S^r_{mn}$$, and

the matrix T^A has an mn component determined by

$T^A_{mn} = T_{mn}$, $U_{mn} > A$,

$T^A_{mn} = 0$ otherwise,

where

U_{mn} represents an mn ~~the mn~~ component of a ~~of the~~ document ~~or pattern~~ frequency matrix of the set of remaining documents, ~~or pattern set D~~ wherein U_{mm} denotes the number of remaining documents in which the m^{th} term occurs and U_{mn} denotes the number of remaining documents in which the m^{th} and n^{th} terms co-occur; and

A denotes a predetermined threshold.

10. **(currently amended)** A method according to claim 9, further comprising:

determining a modified common co-occurrence matrix Q^A on the basis of T^A ; and

in step (b-2), obtaining the document commonality of each remaining document to the set of the remaining documents based on the co-occurrence matrix S^r of each remaining document and the modified common co-occurrence matrix Q^A;

the matrix Q^A having an mn component determined by

$Q^A_{mn} = \log T^A_{mn}$ $T^A_{mn} > 1$,

$Q^A_{mn} = 0$ otherwise.

11. **(currently amended)** A method according to claim 10, wherein in step (b-2),

~~z_{mm} and z_{nn} are respectively weights for a term or object feature m and a term or object feature pair m, n, and~~

~~a.~~ the document ~~or pattern~~ commonality of each remaining document ~~or pattern~~ P having a co-occurrence matrix S^P with respect to the set of remaining documents ~~or pattern set D~~ is given by

~~$com_1(D, P; Q^A) = \frac{\sum_{m=1}^{M} z_{mm} Q^A{}_{mm} S^P{}_{mm}}{\sqrt{\sum_{m=1}^{M} z_{mm} (Q^A{}_{mm})^2} \sqrt{\sum_{m=1}^{M} z_{mm} (S^P{}_{mm})^2}}$ (3)~~

~~or~~

~~$com_q(D, P; Q^A) = \frac{\sum_{m=1}^{M} \sum_{n=1}^{M} z_{mn} Q^A{}_{mn} S^P{}_{mn}}{\sqrt{\sum_{m=1}^{M} \sum_{n=1}^{M} z_{mn} (Q^A{}_{mn})^2} \sqrt{\sum_{m=1}^{M} \sum_{n=1}^{M} z_{mn} (S^P{}_{mn})^2}}$ (4)~~

$$com_q(D', P; Q^A) = \frac{\sum_{m=1}^{M} \sum_{n=1}^{M} Q^A{}_{mn} S^P{}_{mn}}{\sqrt{\sum_{m=1}^{M} \sum_{n=1}^{M} (Q^A{}_{mn})^2} \sqrt{\sum_{m=1}^{M} \sum_{n=1}^{M} (S^P{}_{mn})^2}}.$$

12. **(currently amended)** A method according to ~~claim 9~~ claim 10, wherein in step (b-2),

~~z_{mm} and z_{nn} are respectively weights for a term or object feature m and a term or object feature pair m, n, and~~

~~a.~~ the document ~~or pattern~~ commonality of each remaining document ~~or pattern~~ P having a co-occurrence matrix S^P with respect to the set of remaining documents ~~or pattern set D~~ is given by

~~$com_q(D,P;T^A) = \frac{\sum_{m=1}^{M} z_{mn} T^A{}_{mn} S^P{}_{mn}}{\sqrt{\sum_{m=1}^{M} z_{mn} (T^A{}_{mn})^2} \sqrt{\sum_{m=1}^{M} z_{mn} (S^P{}_{mn})^2}}$~~ ~~(3)~~

~~or~~

~~$com_q(D,P;\boldsymbol{T}^A) = \frac{\sum_{m=1}^{M} \sum_{n=1}^{M} z_{mn} T^A{}_{mn} S^P{}_{mn}}{\sqrt{\sum_{m=1}^{M} \sum_{n=1}^{M} z_{mn} (T^A{}_{mn})^2} \sqrt{\sum_{m=1}^{M} \sum_{n=1}^{M} z_{mn} (S^P{}_{mn})^2}}$~~ ~~(4)~~

$$com_q(D', P; \boldsymbol{T}^A) = \frac{\sum_{m=1}^{M} \sum_{n=1}^{M} T^A{}_{mn} S^P{}_{mn}}{\sqrt{\sum_{m=1}^{M} \sum_{n=1}^{M} (T^A{}_{mn})^2} \sqrt{\sum_{m=1}^{M} \sum_{n=1}^{M} (S^P{}_{mn})^2}}.$$

13-22. (canceled)

23. (Original) A computer arranged to perform the method of claim 1.

24. (Original) A computer arranged to perform the method of claim 2.

25-26. (canceled)

27. (Original) A computer arranged to perform the method of claim 5.

28. (Original) A computer arranged to perform the method of claim 6.

29. **(currently amended)** A clustering apparatus for clustering documents ~~or patterns~~ each having one or plural document ~~or pattern~~ segments in an input document ~~or pattern~~ set, the apparatus comprising:

a first unit for obtaining a co-occurrence matrix for each input document, and obtaining an input document ~~or pattern~~ frequency matrix for the set of input documents ~~or patterns~~, based on occurrence frequencies of terms or term pairs appearing in the set of input documents ~~or pattern~~;

a second unit for selecting a seed document ~~or pattern~~ from a set of remaining documents ~~or patterns~~ that are not included in any cluster existing at that moment and constructing a current

cluster of an initial state based on ~~using~~ the seed document, said second unit being configured for ~~or pattern, wherein said selecting comprises~~

constructing a remaining document common co-occurrence matrix for the set of the remaining documents based on a product of corresponding components of the co-occurrence matrices of all documents in the set of remaining documents; ~~or patterns; and~~

obtaining a document commonality of each remaining document to the set of the remaining documents based on a product sum between every component of the co-occurrence matrix of each remaining document and the corresponding component of the remaining document common co-occurrence matrix;

~~using the common co-occurrence matrix to~~ extracting, as the seed document ~~or pattern~~, the document ~~or pattern~~ having the highest document ~~or pattern~~ commonality to the set of the remaining documents ~~or patterns~~; and

constructing the initial cluster by including the seed document and neighbor documents similar to the seed document;

a third unit

~~for obtaining the document or pattern commonality to the current cluster for each document or pattern in the input document or pattern set using information based on the document or pattern frequency matrix for the input document or pattern set, information based on the document or pattern frequency matrix for documents or patterns in the current cluster and information based on a common co-occurrence matrix of the current cluster, and~~

for making documents, which have ~~or patterns having~~ the document ~~or pattern~~ commonality to the current cluster higher than a threshold, belong temporarily to the current cluster; wherein said third unit is configured for:

constructing a current cluster common co-occurrence matrix for the current cluster and a current cluster document frequency matrix of the current cluster based on occurrence frequencies of terms or term pairs appearing in the documents of the current cluster;

obtaining a distinctiveness value of each term and each term pair for the current cluster by comparing the input document frequency matrix with the current cluster document frequency matrix;

obtaining weights of each term and each term pair from their distinctiveness values;

obtaining a document commonality to the current cluster for each document in the input document set based on a product sum between every component of the co-occurrence matrix of the input document and the corresponding component of the current cluster common co-occurrence matrix while applying the respective weights to said components; and

making documents having the document commonality to the current cluster higher than the threshold belong temporarily to the current cluster;

a fourth unit for repeating the operations of the third unit until the number of documents ~~or patterns~~ temporarily belonging to the current cluster ~~becomes the same as that in the previous repetition~~ does not increase;

a fifth unit for repeating the operations of the second through fourth units until given convergence conditions are satisfied; and

a sixth unit for deciding, on the basis of the document ~~or pattern~~ commonality of each document ~~or pattern~~ to each cluster, a cluster to which each document ~~or pattern~~ belongs, and for outputting said cluster.

30. **(currently amended)** A clustering apparatus according to claim 29, wherein the remaining document common co-occurrence matrix or the current cluster common co-occurrence matrix reflects co-occurrence frequencies at which pairs of different terms co-occur in each document ~~or pattern~~ of the remaining documents or ~~patterns~~ the current cluster, respectively.

31. **(currently amended)** A method according to claim 1, wherein the remaining document common co-occurrence matrix or the current cluster common co-occurrence matrix reflects co-occurrence frequencies at which pairs of different terms co-occur in each document ~~or pattern~~ of the remaining documents or ~~patterns~~ the current cluster, respectively.